

EBM, HTA, and CER: Clearing the Confusion

BRYAN R. LUCE, MICHAEL DRUMMOND,
BENGT JÖNSSON, PETER J. NEUMANN,
J. SANFORD SCHWARTZ, UWE SIEBERT,
and SEAN D. SULLIVAN

The International Working Group for HTA Advancement; United BioSource Corporation; University of York; Stockholm School of Economics; Tufts Medical Center; University of Pennsylvania; University for Health Sciences, Medical Informatics and Technology; University of Washington

Context: The terms *evidence-based medicine* (EBM), *health technology assessment* (HTA), *comparative effectiveness research* (CER), and other related terms lack clarity and so could lead to miscommunication, confusion, and poor decision making. The objective of this article is to clarify their definitions and the relationships among key terms and concepts.

Methods: This article used the relevant methods and policy literature as well as the websites of organizations engaged in evidence-based activities to develop a framework to explain the relationships among the terms EBM, HTA, and CER.

Findings: This article proposes an organizing framework and presents a graphic demonstrating the differences and relationships among these terms and concepts.

Conclusions: More specific terminology and concepts are necessary for an informed and clear public policy debate. They are even more important to inform decision making at all levels and to engender more accountability by the organizations and individuals responsible for these decisions.

Keywords: Evidence, policy, decision making, evidence-based medicine, health technology assessment, comparative effectiveness research.

Address correspondence to: Bryan R. Luce, Center for Health Economics and Science Policy, United BioSource Corporation, 7101 Wisconsin Avenue, Suite 600, Bethesda, MD 20814 (email: bryan.luce@unitedbiosource.com).

MANY JURISDICTIONS ARE BECOMING INCREASINGLY interested in evidence-based health care decision making because of their desire to improve the quality and efficiency of care provided to patients. Such activities are given various names, such as *evidence-based medicine* (EBM), *health technology assessment* (HTA), or more recently, *comparative effectiveness research* (CER). These terms are not used consistently, however, which has led to confusion in the medical and health policy communities. In this article, we propose an organizing framework that relates these evidence terms to three basic questions regarding evidence generation, evidence synthesis, and decision making: Can it work? Does it work? Is it worth it?

The lack of consistency and clarity, and even the misuse of basic words and terms of evidence-based activities, leads to unnecessary disagreements among key stakeholders concerning their appropriate role in health care decision making. By “key stakeholders,” we mean medical professionals and other health care providers, health care payers, legislators and regulators, private and public health agencies, manufacturers, patients, patient advocacy groups, and, finally, taxpayers. The implications can be significant because the evidence processes to which we refer include the development and application of clinical guidelines, insurance coverage decisions, drug formulary placement, and reimbursement decisions, that is, the payment for and access to health care itself. When EBM, HTA, or CER is invoked, what is meant by the term *evidence*? Is it limited to evidence from randomized clinical trials? Or does it also include evidence from well-designed observational studies? Which elements of evidence are in and which are out? What about evidence of costs to health plans or patients’ out-of-pocket costs? How about lost or gained productivity? Patients’ preferences? Or health-related quality of life?

Among the many examples of this problem that we address here is that of Dr. David Sackett, who popularized the concept and application of EBM (Sackett et al. 2000): “The practice of evidence-based medicine means integrating individual clinical expertise with the best available external clinical evidence from systematic research . . . and individual patients’ predicaments, rights and preferences in making clinical decisions about their care” (Sackett et al. 1996, 71). This definition is very different from Eddy’s (1997, 2005), as he adopts a broader definition of EBM that also considers the development of evidence-based policies and guidelines, the view taken as well by the Institute of Medicine

Roundtable on Evidence-Based Medicine (IOM 2009). According to Eddy, the broader notion of EBM also encompasses cost-effectiveness.

As a concept, “health technology assessment” was initially established by the Office of Technology Assessment in 1978 (Office of Technology Assessment 1978) and continues to be defined by the International Network of Agencies for Health Technology Assessment (INAHTA) as “a multidisciplinary field of policy analysis, studying the medical, economic, social and ethical implications of development, diffusion and use of health technology” (INAHTA 2009). In this context, “technology” is defined broadly to include drugs, devices, procedures, and systems of organization of health care, although in practice, it is commonly applied much more narrowly. In Canada, Europe, and many other parts of the world, agencies designated as HTA organizations have social legitimacy; they continue to flourish and the assessments almost always include cost-effectiveness. The United States, however, has shifted away from using the term *HTA* and seems to have substituted terms like *outcomes research*, *effectiveness research*, *EBM*, and, most recently, *comparative effectiveness research*. Unfortunately, the sequential substitution of these terms for HTA has not been accompanied by clarity of meaning, other than it usually does not include economic evaluation in the United States (Luce and Cohen 2009).

The term *CER* also is used differently by different individuals and organizations. One example is a recent paper (Chalkidou et al. 2009) that reviewed the CER experiences of agencies in three European countries that undertook a range of work, including activities that have traditionally been termed HTA. In the United States, where CER became a focal issue in the 2009/2010 health reform debate, its definition has taken on multiple dimensions and meanings and has not been clearly differentiated from either HTA or EBM. A typical definition of CER is that of the Institute of Medicine (IOM):

CER is the generation and synthesis of evidence that compares the benefits and harms of alternative methods to prevent, diagnose, treat and monitor a clinical condition, or to improve the delivery of care. The purpose of CER is to assist consumers, clinicians, purchasers, and policy makers to make informed decisions that will improve health care at both the individual and population levels. (IOM 2009)

This definition is sufficiently broad to mean entirely different things to different stakeholders and policymakers. Some would consider CER to

be a single head-to-head comparative trial or other single comparator study, and others would consider it a synthesis of existing evidence of alternative interventions. To most Europeans, the IOM's definition refers to HTA and includes economic evaluation.

Some analysts in the United States, most notably Gail Wilensky, who sparked the call in 2006 for a major national commitment to CER (Wilensky 2006), speak of it largely in regard to investing in comparative effectiveness trials (Wilensky, personal communication, 2010). Many others, however, seem to consider CER mainly as an extension of and a deeper commitment to improving comparative evidence using traditional health services research methods, for example, via literature synthesis, observational methods, and improved health information technology. Some analysts advocate including economic evaluation in CER (American College of Physicians 2008), while others want to exclude it (Wilensky 2008) or have not commented on its inclusion or exclusion (Federal Coordinating Council 2010; IOM 2009).

More precise terminology will not be sufficient to ensure wise investment in improving the evidence base and the optimal use of the evidence. But it is necessary not only to enable public policy debate to be informed and clear but also, and more important, to improve decision making at all levels and to make the organizations and individuals responsible for these decisions more accountable.

This is the objective of our article.

We begin by proposing a general organizational framework and then depict in some detail our concept of the interrelationships of EBM, HTA, CER, and related terminology.

An Organizing Framework

In order to better illustrate the current confusion and to help derive more precise definitions of EBM, HTA, CER, and related concepts, we propose a three-by-three matrix. Along one axis are the three questions that evidence-based processes in health care seek to answer about an intervention, namely, "Can it work" (i.e., efficacy), "Does it work?" (i.e., effectiveness), and "Is it worth it?" (i.e., economic value). In the case of economic value, we also distinguish between value to the patient and value to the payer/society. Along the other axis are the three key

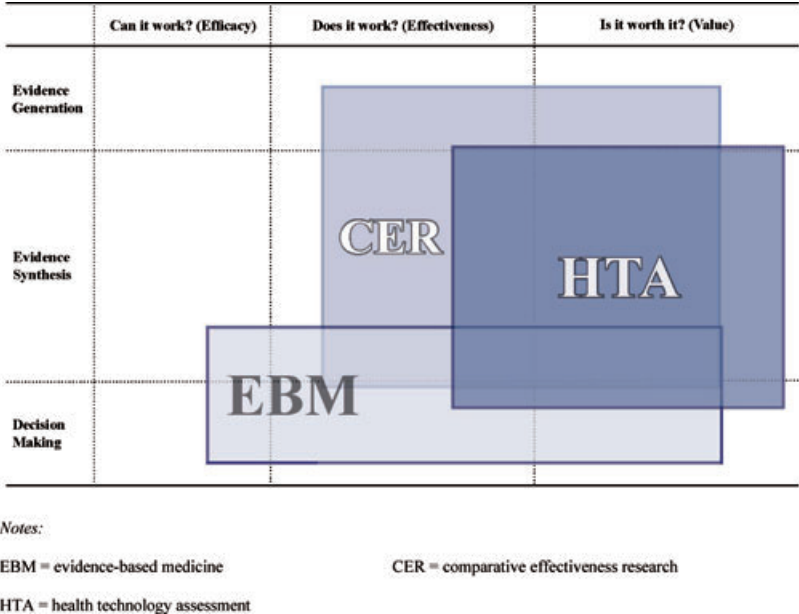


FIGURE 1. Current Confusion over Views of EBM, CER, and HTA.

functions of implementing evidence-based activities, namely, “Evidence Generation,” “Evidence Synthesis,” and “Decision Making.”

As figure 1 shows, if we depicted the current notions of EBM, HTA, and CER as rectangles on the matrix based on definitions implied by comments in the current literature, we would immediately see why there is so much confusion. The space occupied by EBM covers all three evidence questions plus the functions of both evidence synthesis and decision making. The space occupied by HTA is similar to that of EBM, except that it stretches further into the domain of value to society and is perhaps less influenced by efficacy. Although EBM is not depicted as including primary clinical evidence generation, the box for HTA does cross into this space, since in some cases HTA studies may include primary economic evidence generation. The space occupied by CER is depicted as residing mainly in evidence generation and synthesis but also overlapping into decision making, notwithstanding that some U.S. commentators (e.g., Wilensky 2008) and legislation (e.g., U.S. Congress 2003) have made clear that the process of making coverage and other policy-related decisions should or will remain separate.

This particular depiction of confused but related concepts may be somewhat U.S. centric, but we would argue that our general assertion of confusion exists outside the United States, too.

More Precise Definitions

To define the three key concepts more precisely, we need to describe in more detail the contents of each of the cells. We discuss how evidence is generated, synthesized, and used in decision processes for each of the three evidence questions. In the medical field, evidence of safety and efficacy is generated primarily from clinical studies of various designs, as described later. For decision-making purposes, however, all the available evidence on a given topic needs to be synthesized.

Efficacy: “Can It Work?”

A health care intervention is considered efficacious when there is evidence that the intervention provides the intended health benefit when administered to carefully selected patients according to prescribed criteria, often by experts in a research setting. The evidence of efficacy typically comes from well-controlled randomized clinical trials (RCTs), as typically conducted on new medicines. Because RCTs use randomization and other research design features that minimize potential biases (e.g., confounding, information, selection bias), they are essential to help demonstrate a causal relationship between an intervention and an outcome (i.e., they have high internal validity). Efficacy/RCT trials are commonly—and appropriately—described as “explanatory” trials. That is, although they are often designed to evaluate “Can it work,” they also are often useful in explaining *why* it works (Schwartz and Lellouch 1967/2009). These trials are usually designed to answer precisely quite narrow clinical or biological questions.

Evidence synthesis relevant to the “Can it work” question is conducted by a systematic review of trials (SRT), in which a thorough (systematic) search of the trial literature is conducted and summary estimates of the key clinical parameters are produced, often through a formal process such as meta-analysis. The main and most appropriate decision process using efficacy RCTs is that of licensing/market approval for health technologies, as carried out by the Food and Drug Administration (FDA) in

the United States and similar bodies in other jurisdictions, that require evidence of efficacy, as well as the safety and quality of manufacture. These trials also often form the basis for clinical guidelines and clinical decisions, although as noted in the next section, guidelines and decisions should be also informed by the “Does it work” question.

Evidence of efficacy from RCTs is appropriately considered to be the gold standard study design in the evidence hierarchy, but only for the “Can it work?” question. Conversely, it is not necessarily the gold standard for the “Does it work?” question, the subject of the next section. Unfortunately, as we point out later, efficacy trials are all too commonly misspecified as leading to evidence of effectiveness, which can result in overinterpreting their value and relevance to routine practice settings and the resulting decisions.

Effectiveness: “Does It Work?”

A health care intervention is considered effective when there is evidence of benefit to patients when administered by physicians in routine clinical practice settings. Factors such as surgeons’ learning curve with a new procedure or patients’ adherence to drug therapies can cause the effectiveness of a health care intervention to differ substantially from its efficacy. By its very nature, a study that produces evidence of effectiveness in real-world practice settings is more generalizable. It tends to trade aspects of internal validity for higher external validity and typically addresses broader questions than do efficacy trials.

RCTs that are conducted in routine practice settings are called *effectiveness RCTs* or, more commonly, *pragmatic* or *practical clinical trials* (PCTs) (Schwartz and Lellouch 1967/2009; Tunis, Stryer, and Clancy 2003). As there is no hard-and-fast distinction between efficacy and pragmatic trials, it is probably better to think of a spectrum of trial designs and conduct (Thorpe et al. 2009). It is also true that some RCTs are conducted in routine practice settings, a trend that we hope continues. The distinguishing features of pragmatic trials are that they are always conducted in routine practice settings; they normally compare the new intervention with an accepted current practice; and they tend to be more focused on end points of direct interest to the patient (e.g., functional status, final health outcome), as opposed to surrogate end points such as viral load or blood pressure. In addition, the protocols for such trials

are usually much less intrusive (i.e., fewer prescribed visits, fewer measurements of clinical parameters, and less patient monitoring), and the actual rates of adherence to therapy may be of study interest, rather than adherence being specifically encouraged, as is typical in an efficacy RCT. For instance, in a pragmatic trial comparing the newer, more expensive antidepressant (fluoxetine-Prozac) with inexpensive generic alternatives (imipramine and desipramine), the “pragmatic” research question was “What will happen to the health of our patients and to our plan’s budget if we initiate treatment with drug X or Y in the routine primary care setting?” (Simon et al. 1996). That is, almost the only clinical protocol feature was initial random assignment to therapy. All patients whom the primary physician deemed qualified for an antidepressant drug were eligible, and clinical decisions and patients’ behavior (e.g., adherence to therapy) were completely naturalistic.

In other situations, a clinical trial cannot be conducted because of logistical or ethical reasons, or a trial may not be the most efficient way of gathering the data. The latter is often the case for rare side effects or outcomes in the longer term. In such cases, it often is necessary to use uncontrolled studies such as registries, administrative databases, or clinical case series. Such studies are used by bodies like the FDA to assess safety, but these observational studies can also be important sources of evidence for effectiveness, even though they do not have the high internal validity of RCTs.

One reason this distinction between efficacy and effectiveness is important is that the misapplication of the terms may be contributing to a false sense of what constitutes “best synthesized evidence” for informed decision making about the effects of health interventions in routine practice settings. Decisions based on such misconceptions may adversely affect patients’ access to care and innovative products, the quality of care, and the costs and cost-effectiveness of delivering such care. As noted by Lohr and colleagues, “The fundamental basic science imperative may be to generate information about the efficacy of health care interventions, but the practical realities of policy and economic decisions call for knowledge about effectiveness” (Lohr, Eleazer, and Mauskopf 1998, 16).

Therefore, synthesis of effectiveness data, which we term *systematic review of evidence* (SRE), typically includes data from well-conducted observational studies as well as RCTs. As a rule, efficacy RCT data and real-world effectiveness data complement one another and are best used in tandem. Although the literature search for SRE is, or should be,

as rigorous as that in SRTs, the inclusion criteria for studies are less restrictive, and the methods of summarizing evidence on key clinical end points are typically, and often necessarily, less formal in order to address a broader range of questions.

Decision processes relating to effectiveness evidence reviews are highly varied and may include the development of clinical guidelines (e.g., as part of EBM), individual patient-physician care decisions, or inputs to the decisions about economic value, which we address later. Here, we generally place decisions about economic value in the HTA category.

Several of the world's most prominent evidence-based medicine-related organizations have relied almost exclusively on evidence of health care interventions derived from randomized clinical trials (Center for Evidence-Based Policy 2010; Cochrane Collaboration 2010). These studies are primarily efficacy RCTs, particularly in the case of new medicines for which trials are conducted mainly for licensing purposes, although occasionally the evidence base may include trials that are closer to the pragmatic end of the spectrum. A heavy reliance on efficacy RCTs is not a major problem in the work of a registration authority such as the FDA, since its role is to establish whether a new technology can work for the purpose of market approval.

The principal problem arises when organizations seeking to inform real-world choices of therapies rely solely on efficacy RCTs to inform the "Does it work?" question. Here, it is important to know how interventions work in practice, as opposed to whether they can work in principle. At the international level, organizations seeking to inform such real-world choices include Germany's Institute for Quality and Efficiency in Health Care (IQWiG 2009a), Common Drug Review at the Canadian Agency for Drugs and Technologies in Health (CADTH) (CADTH 2009), and the United Kingdom's National Institute for Health and Clinical Excellence (NICE) (NICE 2009). Within the United States, they include the Drug Effectiveness Review Project (DERP) at the Oregon Health and Science University (Center for Evidence-Based Policy 2010), the Medicare Evidence Development and Coverage Advisory Committee (MEDCAC) (CMS 2009b), and Blue Cross Blue Shield's Technology Evaluation Center (TEC) (Blue Cross Blue Shield 2009).

It is important that these organizations consider a wider range of available, germane evidence, including PCTs and observational data, because the gold standard evidence base for these decisions would be one that emphasized studies conducted in routine practice settings or,

as Rawlins (2008) would say, are “fit for purpose,” taking into account both external and internal validity. A recent study (Neumann et al. 2010) evaluated a number of HTA organizations around the world as to whether they conformed to a set of published principles (Drummond et al. 2008), one of which addresses the inclusion of all relevant evidence. Within the United States, several managed care entities have adopted the Academy of Managed Care Pharmacy’s (AMCP) Format for Formulary Submission (AMCP 2009), which encourages the consideration of a broad range of evidence. The same applies to the methodology guidelines developed by NICE in the United Kingdom (NICE 2008).

Distinguishing efficacy from effectiveness and emphasizing its importance to decision making dates back to at least 1978 (Office of Technology Assessment 1978), so it is far from a new concept. But it is somewhat disconcerting that confusion still exists. Some of this confusion can be traced to the FDA’s authorizing legislation and regulations in which the terms *efficacy* and *effectiveness* are often used interchangeably, so *effectiveness* is used when *efficacy* is intended.¹ This misapplication of terms has carried over to other organizations as well. For example, the Drug Effectiveness Review Project’s stated mission is to “obtain the best available evidence on *effectiveness* [italics added] and safety comparisons between drugs in the same class, and to apply the information to public policy and related activities” (Center for Evidence-Based Policy 2010), yet DERP has tended to rely exclusively on evaluations based on RCTs. Similarly, the highly regarded Cochrane Collaboration describes its reviews as exploring “the evidence for and against the *effectiveness* [italics added] and appropriateness of treatments . . . in specific circumstances” (Cochrane Collaboration 2009), though even a casual examination of the Cochrane reviews indicates a nearly complete reliance on RCT literature. The same is true of some of the German IQWiG’s early benefit assessments, such as that on the insulin analogues (IQWiG 2009b). Based on these arguments, these bodies are relying mainly on evidence of efficacy, as opposed to effectiveness, unless the body of RCT evidence includes a substantial proportion of trials adopting a pragmatic methodology.

Economic Value: “Is It Worth It?”

For a health care intervention to be “worth it,” the value of the clinical and economic benefits needs to be greater than the clinical harms and

economic costs, since the resources that the treatment consumes are not available for other health care purposes. Although the term *value* is commonly used both in and out of health care, it has not been defined in a way that resonates well within the health care evidence community. Value is usually said to be in the eyes of the beholder, which is a valid way to consider the term, although it is not helpful for policymaking purposes. However, we distinguish between value to the patient and value to the payer/society, since these may be different.

The principal way to determine whether the benefits obtained from a given intervention justify the economic costs and potential clinical harms is to compare an economic evaluation of a therapy with relevant alternatives in the setting concerned. Although economic evaluations may include generating evidence, they frequently consist of synthesizing the available clinical and economic data. Since the objective is to assess costs and benefits in routine practice settings, economic evaluations use effectiveness, as opposed to efficacy, data, once the latter (efficacy) is established and when the former (effectiveness) is available. Economic evaluations often are conducted as part of a coverage or reimbursement process or, for our purposes here, within an HTA that in turn may inform a coverage or reimbursement process. In such cases, the value question is typically approached from an aggregate perspective, for instance, that of a health plan or a national health budget. Also of note is that some organizations, such as the U.S. Centers for Disease Control and Prevention's (CDC's) Advisory Committee on Immunization Practices (ACIP), include cost-effectiveness considerations in their clinical guidelines (Smith, Snider, and Pickering 2009).

Requirements to generate new evidence are now beginning to be embedded in coverage schemes. From the payer's/society's perspective, the coverage decision for a technology may be conditioned on further evidence regarding the comparative effectiveness and possible costs of the health care intervention concerned (Levin et al. 2007). Claxton, Sculpher, and Drummond (2002) refer to this as *conditional reimbursement*; in Sweden, the term *restricted reimbursement* is used (TLV 2009). The U.S. Centers for Medicare and Medicaid Services (CMS) has coined the term *coverage with evidence development* (CED) (CMS 2009a), but in keeping with the current Medicare coverage policy, the CED process does not include economic evaluation. Other countries have roughly analogous processes, for instance, the United Kingdom has a formal conditional coverage-like process termed *Only in Research* (see, e.g., Chalkidou et al., 2009).

An element of economic value also can be (and, we argue, should be) contained in the EBM decision process specifically as it relates to patients' out-of-pocket exposure such as copay levels. Although seldom considered, whether in clinical guidelines or literature addressing EBM processes, patients' preferences are undoubtedly based at least in part on out-of-pocket costs and sometimes may affect their personal decisions about alternative therapies and, perhaps most important, affect their patterns of adherence.

Redefining EBM, HTA, and CER

Armed with the insights provided by this organizing framework, we propose clarifying the typology, nomenclature, and interrelationships of the evidence terms discussed in this article. Figure 2 depicts EBM, HTA, and CER again as rectangles. The key related concepts are in circles and ovals, and the most important decision-making processes are shown as diamonds. The arrows illustrate the principal relationships among the various concepts. The dotted lines labeled A through C indicate relationships about which there is considerable dispute, as we discuss later.

Considering first the column headed "Can it work?" we can see that evidence generated from traditional efficacy RCTs is used for decisions about the market approval of new interventions. These RCTs are also currently the major input to systematic reviews of trials. SRTs include, too, pragmatic clinical trials that are becoming more common, and they also are inputs to clinical guidelines, which we view as straddling the first two columns in the diagram.

Comparative-effectiveness research is firmly situated in the column "Does it work?" covering both evidence generation and evidence synthesis. That is, we expect a two-way link between CER and pragmatic clinical trials, and we expect a similar link between CER and SREs. In turn, SREs (systematic reviews of evidence) consider evidence from RCTs, PCTs, and observational studies. We also depict the outputs of CER activities as influencing EBM either directly or indirectly through clinical guidelines, whose development we consider to be one activity of EBM. The outputs of CER activities also are an important input to HTA activities (e.g., the results of nationally funded CER studies may be an important input to HTAs conducted by different payers). One important aspect of this definition of CER is that it has no direct link to any decision process, although its purpose is to do so and thus it ideally

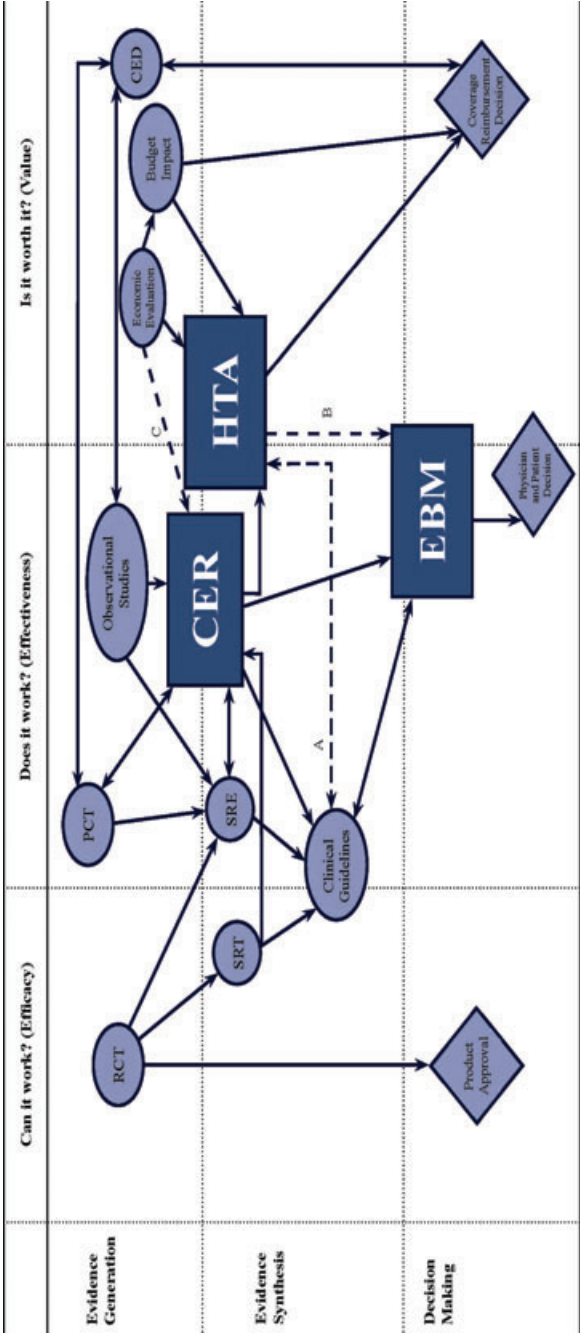


FIGURE 2. Redefined Relationships of Evidence Processes.

influences patients' and physicians' decisions through EBM, or coverage decisions including those through HTA.

EBM is characterized as a decision process, focusing on decisions by individual patients and physicians, but it does cross into the "evidence synthesis" space. In our representation, EBM focuses mainly on the question "Does it work?" although it also contains an element of the question "Is it worth it?" specifically from the patient's perspective. As mentioned previously, we acknowledge that the development of clinical guidelines is an important aspect of EBM, but we think of it from a patient's perspective rather than a societal perspective. We regard the production of guidelines from a societal perspective as being closer to HTA.

Health technology assessment straddles the last two columns in the figure, and is viewed as a method of evidence synthesis that receives inputs from CER, economic evaluation, and the consideration of social, ethical, and legal aspects. As we depict it, HTA is a main input to coverage decisions, which may also be influenced by budget implications.

Finally, conditional coverage (depicted in the diagram as CED) is represented as both a decision and an evidence generation process involving the commissioning and use of observational studies and, occasionally, pragmatic RCTs.

Although we believe that our organizing framework and accompanying definitions do help distinguish concepts and depict relationships of EBM, HTA, and CER (and other related terms), some relationships remain the subject of debate. These are identified by the broken lines in the figure. Line A indicates that there could be a link between clinical guidelines and HTA, opening up the possibility that clinical guidelines may—and sometimes do—explicitly include a consideration of cost-effectiveness and, through HTA, be linked to coverage decisions. We consider this to be entirely appropriate, and it is the approach followed by NICE in the United Kingdom and other countries (see, e.g., *Value in Health* 2009). Our diagram reflects the current majority view that clinical guidelines, as practiced through EBM, are primarily concerned with improving the quality of care for the patient rather than increasing value for money for the payer or society.

Line B suggests that there could be a link between EBM and HTA. First, as we acknowledged earlier, some health technology assessments may focus on the clinical effectiveness and quality of care, rather than

on value for money. Second, as we argued earlier, some of the objectives of quality of care and value for money are closely aligned. For example, removing interventions that contribute only little to improving patients' outcomes and are inconvenient to the patient, are expensive, or are risky will increase the quality of care and save resources. Nevertheless, we still feel that it is preferable to distinguish between activities that concentrate primarily on benefits to the patient (EBM) and those that concentrate primarily on benefits to society at large (HTA).

Line C, going directly from Economic Evaluation to CER, is probably the most controversial of all. This relationship implies that CER studies should include consideration of cost-effectiveness, a suggestion that has led to considerable debate in the United States (American College of Physicians 2008; Wilensky 2008). At the time of this writing, however, U.S. legislation, for instance, that associated with health care reform, does not include cost-effectiveness (U.S. Congress, Senate 2009). We can see the merits of both sets of arguments. But our basic diagram, without line C, has balance, in that CER studies are viewed as an important source of information for those conducting both EBM and HTA.

Those who believe that cost-effectiveness should be routinely incorporated into CER make two arguments. First, CER studies represent an important opportunity for collecting data on resource use and cost, which in turn are useful for subsequent HTAs. Second, if cost-effectiveness is not part of CER, subsequent HTAs may not be fully informed. For example, individual health plans may not have the expertise, time, resources, or even inclination to conduct proper cost-effectiveness analysis. In such cases, the argument is that coverage decisions tend to focus on the acquisition cost of a new technology, rather than to consider fully its economic value.

The counterargument is that the direct inclusion of cost-effectiveness considerations in CER studies changes the overall balance in the emphasis and use of these studies, concentrating on their use in HTA and possibly disregarding their use in EBM. This may lead to objections that would hamper the overall CER movement. We firmly believe, though, that because all decisions have resource implications, cost-effectiveness considerations should play a role in coverage decisions, through either inclusion in CER studies directly or adequately conducted HTAs.

Finally, we note the potential for inherent conflicts that inevitably resides in some, if not all, the concepts that we have tried to clarify and distinguish from one another in this article. For example, consider EBM

and HTA. If we take as given that the proper and full expression of EBM includes individual patients' values, which itself can include patients' out-of-pocket costs, the societal aggregate of these values would quickly conflict with societal HTA. This is because the former assumes that individual patients are to be satisfied one by one, whereas HTA may take a societal approach, often within budget constraints, meaning that not everyone may get everything he or she desires. Thus, although we seek to differentiate, bind, and relate EBM-HTA-CER (plus other related evidence-based processes) within one holistic graphic, we are not contending that they all have a central unifying aspect.

Bearing all this in mind, the following are our preferred definitions of the three key terms:

1. Evidence-based medicine (EBM) is an evidence synthesis and decision process used to assist patients' and/or physicians' decisions. It considers evidence regarding the effectiveness of interventions and patients' values and is mainly concerned with individual patients' decisions, but is also useful for developing clinical guidelines as they pertain to individual patients.
2. Comparative effectiveness research (CER) includes both evidence generation and evidence synthesis. It is concerned with the comparative assessment of interventions in routine practice settings. The outputs of CER activities are useful for clinical guideline development, evidence-based medicine, and the broader social and economic assessment of health technologies (i.e., HTA).
3. Health technology assessment (HTA) is a method of evidence synthesis that considers evidence regarding clinical effectiveness, safety, cost-effectiveness and, when broadly applied, includes social, ethical, and legal aspects of the use of health technologies. The precise balance of these inputs depends on the purpose of each individual HTA. A major use of HTAs is in informing reimbursement and coverage decisions, in which case HTAs should include benefit-harm assessment and economic evaluation.

Conclusions

Evidence-based processes for health care decision making have attracted increased interest. While this is a welcome development, its

potential impact may not be fully realized owing to confusion over terms such as *evidence-based medicine*, *health technology assessment*, *comparative effectiveness research*, and other related terms. We believe that more precise terminology will benefit the public policy dialogue, general health and clinical policy decision making, and specific clinical decisions by patients and physicians. It will also promote accountability by the organizations and individuals responsible for these decisions and ultimately enhance the quality of patient care.

In this article, we offer more precise definitions of key terms based on an organizing framework that clarifies the differences and relationships among EBM, HTA, CER, and related concepts. All three terms address the “Does it work?” question, but none asks the “Can it work?” question. Health technology assessment is the primary activity that considers “Is it worth it?” although as we point out, EBM should also address the more limited question “Is it worth it to the patient?” taking into account the costs to the patient. Of the three concepts, only EBM is a decision process, although both HTA and CER are applied specifically to feed into decision making.

Endnote

1. For example, see “Guidance for Industry: Providing Clinical Evidence of Effectiveness for Human Drugs and Biological Products,” available at <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM078749.pdf> (accessed November 24, 2009). In this document, note 2 states: “As used in this guidance, the term efficacy refers to the findings in an adequate and well-controlled clinical trial . . . and the term effectiveness refers to the regulatory determination that is made on the basis of clinical efficacy and other data.” However, it is well understood that FDA decisions are based nearly solely on safety and efficacy. This is seen in the FDA’s documents referring to regulatory approval resting on meeting the “substantial evidence” requirement. “Substantial evidence” was defined in section 505(d) of the Food, Drug, and Cosmetic Act as “evidence consisting of adequate and well-controlled investigations, including clinical investigations, by experts qualified by scientific training and experience to evaluate the effectiveness of the drug involved, on the basis of which it could fairly and responsibly be concluded by such experts that the drug will have the effect it purports or is represented to have under the conditions of use prescribed, recommended, or suggested in the labeling or proposed labeling thereof.” This, however, is the appropriate definition of “efficacy,” not “effectiveness.”

References

- AMCP (Academy of Managed Care Pharmacy). 2009. The AMCP Format for Formulary Submissions. Version 3.0. Available at http://www.fmcenet.org/index.cfm?c=news.details&a=wn&id=C604C25D_2_1~Final_Final.pdf (accessed November 25, 2009).
- American College of Physicians. 2008. Information on Cost-Effectiveness: An Essential Product of a National Comparative Effectiveness Program. *Annals of Internal Medicine* 148:956–61.
- Blue Cross Blue Shield. 2009. Technology Evaluation Center. Available at <http://www.bcbs.com/blueresources/tec> (accessed November 25, 2009).
- CADTH (Canadian Agency for Drugs and Technologies in Health). 2009. Common Drug Review. Available at <http://www.cadth.ca/index.php/en/cdr> (accessed November 25, 2009).
- Center for Evidence-Based Policy. 2010. The Drug Effectiveness Review Project (DERP). Available at <http://www.ohsu.edu/ohsuedu/research/policycenter/> (accessed February 24, 2010).
- Chalkidou, K., S. Tunis, R. Lopert, L. Rochaix, P.T. Sawicki, M. Nasser, and B. Xerri. 2009. Comparative Effectiveness Research and Evidence-Based Health Policy: Experience from Four Countries. *The Milbank Quarterly* 87(2):339–67. Available at <http://www.milbank.org/quarterly/8702feat.html> (Accessed March 26, 2010).
- Claxton, K., M. Sculpher, and M. Drummond. 2002. A Rational Framework for Decision Making by the National Institute for Clinical Excellence. *Lancet* 360:711–15.
- CMS (Centers for Medicare and Medicaid Services). 2009a. Coverage with Evidence Development. Available at www.cms.hhs.gov/CoverageGenInfo/03_CED.asp (accessed November 23, 2009).
- CMS (Centers for Medicare and Medicaid Services). 2009b. Medicare Evidence Development and Coverage Advisory Committee (MEDCAC). Available at http://www.cms.hhs.gov/FACA/02_MEDCAC.asp (accessed November 25, 2009).
- Cochrane Collaboration. 2009. Cochrane Reviews. Available at <http://www.cochrane.org/reviews/clibintro.htm#intro> (accessed November 25, 2009).
- Cochrane Collaboration. 2010. Available at: <http://www.cochrane.org> (accessed February 25, 2010).
- Drummond, M.F., J.S. Schwartz, B. Jonsson, B. Luce, P. Neumann, U. Siebert, and S. Sullivan. 2008. Key Principles for the Improved Conduct of Health Technology Assessments for Resource Allocation Decisions. *International Journal of Technology Assessment in Health Care* 24(3):244–58.

- Eddy, D.M. 1997. Investigational Treatments: How Strict Should We Be? *Journal of the American Medical Association* 278:179–85.
- Eddy, D.M. 2005. Evidence-Based Medicine: A Unified Approach. *Health Affairs* 24(1):9–17.
- Federal Coordinating Council. 2010. Draft Definition of Comparative Effectiveness Research. Available at <http://www.hhs.gov/recovery/programs/ceer/draftdefinition.html> (accessed March 16, 2010).
- INAHTA (International Network of Agencies for Health Technology Assessment). 2009. HTA Resources. Available at <http://www.inahta.org/HTA/> (accessed November 25, 2009).
- IOM (Institute of Medicine). 2009. *Initial National Priorities for Comparative Effectiveness Research*. Washington, DC: National Academies Press. Available at <http://www.iom.edu/Reports/2009/ComparativeEffectivenessResearchPriorities.aspx> (accessed November 25, 2009).
- IQWiG (Institute for Quality and Efficiency in Healthcare). 2009a. Available at <http://www.iqwig.de/institute-for-quality-and-efficiency-in-health.2.en.html?random=5458e2> (accessed November 25, 2009).
- IQWiG (Institute for Quality and Efficiency in Healthcare). 2009b. *Long-Acting Insulin Analogues in the Treatment of Diabetes Mellitus Type 2*. Cologne. Available at <http://www.iqwig.de/index.558.en.html> (accessed March 26, 2010).
- Levin, L., R. Goeree, N. Sikich, B. Jorgensen, M. Brouwers, T. Easty, and C. Zahn. 2007. Establishing a Comprehensive Continuum from an Evidentiary Base to Policy Development for Health Technologies: The Ontario Experience. *International Journal of Technology Assessment in Health Care* 23:299–309.
- Lohr, K.N., K. Eleazer, and J. Mauskopf. 1998. Health Policy Issues and Applications for Evidence-Based Medicine and Clinical Practice Guidelines. *Health Policy* 46(1):1–19.
- Luce, B.R., and R.S. Cohen. 2009. Health Technology Assessment in the United States. *International Journal of Technology Assessment in Health Care* 25:33–41.
- Neumann, P.J., M.F. Drummond, B. Jonsson, B.R. Luce, J.S. Schwartz, U. Siebert, and S.D. Sullivan. 2010. Are Key Principles for Improved Health Technology Assessment Supported and Used by Health Technology Assessment Organizations? *International Journal of Technology Assessment in Health Care* 26(1):71–78.
- NICE (National Institute for Health and Clinical Excellence). 2008. *Guide to the Methods of Technology Appraisal*. Available at <http://www.nice.org.uk/media/B52/A7/TAMethodsGuideUpdatedJune2008.pdf> (accessed November 30, 2009).

- NICE (National Institute for Health and Clinical Excellence). 2009. Available at www.nice.org.uk (accessed November 25, 2009).
- Office of Technology Assessment. 1978. *Assessing the Efficacy and Safety of Medical Technologies*. September. NTIS order #PB-286929. Available at <http://www.fas.org/ota/reports/7805.pdf> (accessed November 25, 2009).
- Rawlins, M. 2008. *De testimonio*: On the Evidence for Decisions about the Use of Therapeutic Interventions. *Lancet* 372:2152–61.
- Sackett, D.L., W.M. Rosenberg, J.A. Gray, R.B. Haynes, and W.S. Richardson. 1996. Evidence Based Medicine: What It Is, and What It Isn't. *British Medical Journal* 312(7023):71–72.
- Sackett, D.L., S. Straus, S.R. Richardson, W. Rosenberg, J.A. Gray, and R.B. Haynes. 2000. *Evidence-Based Medicine: How to Practice and Teach EBM*. London: Churchill Livingstone.
- Schwartz, D., and J. Lellouch. 1967/2009. Explanatory and Pragmatic Attitudes in Therapeutical Trials. *Journal of Chronic Diseases* 20:637–48. Reprinted in *Journal of Clinical Epidemiology* 62:499–505.
- Simon, G.E., M. VonKorff, J.H. Heiligenstein, D.A. Revicki, L. Grothaus, W. Katon, and E.H. Wagner. 1996. Initial Antidepressant Choice in Primary Care: Effectiveness and Cost of Fluoxetine vs Tricyclic Antidepressants. *Journal of the American Medical Association* 275(24):1897–1902.
- Smith, J.C., D.E. Snider, and L.K. Pickering. 2009. Immunization Policy Development in the United States: The Role of the Advisory Committee on Immunization Practices. *Annals of Internal Medicine* 150(1):45–49.
- Thorpe, K.E., M. Zwarenstein, A.D. Oxman, S. Treweek, C.D. Furberg, D.G. Altman, S. Tunis, E. Bergel, I. Harvey, D. Magid, and K. Chalkidou. 2009. A Pragmatic-Explanatory Continuum Indicator Summary (PRECIS): A Tool to Help Trial Designers. *Journal of Clinical Epidemiology* 62:464–75.
- TLV. 2009. Welcome to TLV. Available at <http://www.tlv.se/in-english/> (accessed November 25, 2009).
- Tunis, S.R., D.B. Stryer, and C.M. Clancy. 2003. Practical Clinical Trials: Increasing the Value of Clinical Research for Decision Making in Clinical and Health Policy. *Journal of the American Medical Association* 290:1624–632.
- U.S. Congress. 2003. *Medicare Prescription Drug, Improvement, and Modernization Act of 2003*. Pub. L. no. 108-173, 117 Stat 2066.
- U.S. Congress, Senate. 2009. *The Patient Protection and Affordable Care Act*. 111th Cong., 1st sess., H.R. 3590. Available at http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=111_cong_bills&docid=f:h3590eas.txt.pdf (accessed March 11 2009).

- Value in Health*. 2009. Special Issue: Health Technology Assessment in Evidence-Based Health Care Reimbursement Decisions around the World: Lessons Learned. 12(s2):S1–S53.
- Wilensky, G.R. 2006. Developing a Center for Comparative Effectiveness Information. *Health Affairs* 25(6):w572–w585.
- Wilensky, G.R. 2008. Cost-Effectiveness Information: Yes, It's Important, but Keep It Separate, Please! *Annals of Internal Medicine* 148:967–68.

Acknowledgments: The International Working Group for HTA Advancement was established in July 2007 with unrestricted funding from the Schering Plough Corporation (now Merck & Co). The Working Group's mission is to provide scientifically based leadership to facilitate significant continuous improvement in the development and implementation of practical, rigorous methods into formal health technology assessment (HTA) systems and processes, by facilitating the development and adoption of high-quality, scientifically driven, objective, and trusted HTA to improve patient outcomes, the health of the public, and the overall quality and efficiency of health care. We are grateful to Emily Sargent for her outstanding graphics work, formatting, and general manuscript preparation and for the many helpful comments by several anonymous reviewers as well as Bradford Gray. The views expressed in this article are those of the authors and do not necessarily reflect the opinions of any of these individuals.